1947

Professional accounting testing program bulletin no. 2; Objective examinations in professional accounting

Leo A. Schmidt

American Institute of Accountants. Committee on Selection of Personnel

THE AMERICAN INSTITUTE OF ACCOUNTANTS
PROFESSIONAL ACCOUNTING TESTING PROGRAM

Bulletin No. 2

OBJECTIVE EXAMINATIONS IN PROFESSIONAL ACCOUNTING,

by Leo. A. Schmidt

Preliminary Edition
For Review and Criticism
(Confidential)

Published by
Committee on Selection of Personnel
437 West 59th Street
New York 19, N. Y.

June, 1947

21 Audubon Ave

32

THE AMERICAN INSTITUTE OF ACCOUNTANTS

PROFESSIONAL ACCOUNTING TESTING PROGRAM

Bulletin No. 2

OBJECTIVE EXAMINATIONS IN PROFESSIONAL ACCOUNTING

by

LEO A. SCHMIDT

Associate Professor

Graduate School of Business Administration

Harvard University

Preliminary Edition (Confidential)

Published by

COMMITTEE ON SELECTION OF PERSONNEL

437 WEST 59TH STREET

NEW YORK 19, N. Y.

JUNE, 1947

CONTENTS

OBJECTIVE EXAMINATIONS IN PROFESSIONAL ACCOUNTING

Introduction

The American Institute of Accountants Committee on Selection of Personnel has embarked upon an ambitious program of testing within the colleges and universities of the country and among the members of the profession in general. The more immediate purpose of this project is to render aid to the members of the profession in their selection, retention and promotion procedures. Backing up this immediate purpose is the longer run purpose of attracting to and retaining in the accounting profession the best possible quality of young men to carry on the profession in the years to come.

The Strong Vocational Interest Test and the Orientation tests developed by the Committee are obviously aimed primarily at the long run guidance problem. The technical accounting achievement tests are concerned with both the functions of guidance and selection for advanced courses or for employment, the emphasis shifting between the several levels at which the tests are to be given.

The first level achievement test is to be given to college students at the end of their first full year course in accounting. This test is aimed primarily at guidance. It is believed that at this point a reliable measure of achievement against nation-wide standards will be, to the young man who shows up well, a powerful incentive toward making accounting his objective. This will be especially true if the Strong test and the Orientation test have already pointed in that direction. Similarly the achievement test should be quite effective in dissuading those who have done poorly from making the unwise choice of going on in accounting. Probably the most important effect of proper winnowing at this stage will be the help it gives to accounting faculties toward putting subsequent accounting courses on a really professional level. This winnowing process has of course always been one of the important functions of the course grades given by the college itself. The standardized accounting tests will thus have a corroborative or reinforcing effect rather than being an entirely new factor. In many instances the effect will be a very welcome aid to the college guidance officers in recommendations which they would otherwise have made with less complete assurance.

The second level of the achievement tests will be given at about the time of college graduation to students who have majored in accounting. At this point the emphasis shifts frankly to the problem of selection for employment. As soon as the tests are taken they will be sent to the office of the Committee, and graded. The grades on a uniform basis for the entire country will be on record in the Committee's office and available on the examinee's authorization to prospective employers designated by the examinee. Each college will receive a report on its own students.

For the benefit of those who will not have access to the level II test through the colleges the Committee is planning public examination centers in the larger cities. The test records of these examinees will be available to prospective employers as are those of the college graduates.

While the test grades thus available should be of considerable aid to the employer in choosing among his applicants for jobs, it is not contemplated that they should be the sole evidence to be considered. In most cases there will also be available the Strong Vocational Interest Blank, as well as complete summaries of the applicant's college record in other studies, extra curricular activities, and part time employment. Most employers will probably continue to require character reference letters, and they certainly should insist on personal interviews. Good employment practice will continue to recognize that the technical skill which these achievement tests measure is only one of the requirements of success in accounting, and that the tremendously important character and personality factors must still be judged by time honored means.

The third level tests of the Committee on Selection of Personnel are not yet as fully developed as the first and second levels. They are intended for use primarily in the accounting firms, to be given to employees who have had three or four years of experience in public accounting and who are being considered for advancement to positions of greater responsibility within the firm. The degree of difficulty will approximate that of the Institute's uniform C.P.A. examination, but the form will still be objective. These tests will attempt to determine the extent to which the examinee has learned the lessons of his several years of experience and to what extent his judgment has developed in matters of public accounting practice.

The Meaning of Objectivity in Examinations

Probably no one would care to dispute the definition of the ideal educational situation so graphically portrayed in Arthur Guiterman's oft quoted poem on Education:

> *For Education is, Making Men;*
> *So is it now, so was it when*
> *Mark Hopkins sat on one end of a log*
> *And James Garfield sat on the other.*

While some of our modern scientists might hold out for a fifty ton cyclotron set up alongside Mark Hopkins' log, even they would be fully appreciative of the great advantage of intimate contact between the teacher and student which the quotation implies is the essence of educational opportunity.

A course of instruction which had proceeded for some weeks or months on the Hopkins - log - student basis might very properly terminate in an examination on the same informal plan. In two or three hours of conversation Mark Hopkins could find out very certainly how well the student had mastered his teachings. Every slight hesitation on the student's part, every inflection of his voice, every gesture, even the lighting up of his eyes at certain moments, would be full of meaning to the examiner who knew him so well.

Sadly, however, education has moved away from the ideal toward the present situation of crowded classes taught under the pressure of numerous distractions by teachers on the average less able and inspiring than Mark Hopkins. Under these conditions examining too must change if it is to be effective. The reliance upon the spoken or written word in conversational or essay style becomes far more hazardous as the intimate acquaintance of the teacher-examiner and the student disappears. The validity of the examiner's subjective insight decreases rapidly in the haze of worsening conditions and his own fallibility. It takes a stronger light than wishful thinking to penetrate the fog of semantic difficulties. Such a stronger light we believe we have in the progressive "objectivisation" of the examination process.

By objectivity in an examination we mean that quality which makes the examinee's answer recognizable as right or wrong, with the exercise of little or no skill or judgment on the part of the grader. This is obviously true of the various short-answer forms, true-false, multiple choice, simple recall, and so forth, in which a circle, check mark, or single word is the entire visible answer. Probably no thoughtful person will quarrel with the desirability of objectivity itself. The real issue is the sacrifice of other essential values that may be consequent upon the attainment of objectivity. A really critical study of this issue is the purpose of this paper.

There is a wide area in examining practice in which objectivity is a relative matter. Analysis of this phase of the subject is of great importance for two reasons. First, an understanding of *relative* objectivity provides the opportunity for substantial improvement of examinations in circumstances where for some reason or another the full objec-

tive form is not advisable. Second, it is through the analysis of relative objectivity that we come most surely to an appreciation of the possibility of breaking down even quite complicated and abstruse matters into the fully objective types of questions.

We are here very close to the crux of the whole philosophy of objective examinations. It is too easy (upon slight acquaintance with the objective technique) to come to the snap conclusion that this technique lends itself well to the examining of "factual matters" (dates, place names, the multiplication table, etc.) but that it cannot be utilized to examine on questions involving reasoning, judgment and imagination, or anything else beyond simple "factual" matters.

Such a conclusion assumes that the subject matter of the examination is inherently objective or non-objective and that nothing can be done about it. This assumption results from a failure to examine the problem closely.

Fundamentally the problem of examining is the same as the problem of studying-learning-teaching. Practical problems in all fields usually come at us in great composite masses. At times in actual life situations we are forced to make decisions about the composite mass with little or no opportunity to break it down into its understandable elements. The systematic learning process, however, characteristically does allow us the opportunity to break the composite mass into its elements, and to assay separately the significance of the separate static facts or independent active forces found therein. Often, in this process, what was thought to be an elemental fact or force at one stage of such an investigation is found to be in itself a complex requiring still further analysis.

At whatever stage of analysis we have arrived, the problem of educational inquiry is to describe the static facts and active forces clearly, for purposes of record and to make possible the communication of our new knowledge to others. The proper description of any significant factor must be in terms of (1) its directional effect and (2) its weight in relation to other factors.

If the above is a correct analysis of the study-learning-teaching process, is it not likely that the examining process may effectively be based on parallel lines?

When the examinee is given a great composite question to discuss, the entire double burden of analytical breakdown and assaying the many factors is in his hands. The perfect student would rise to such a task and produce a complete and correct analysis with full judgments on every factor. The typical student, however, will produce an inadequate analysis; he will fail entirely to consider many factors which should be considered, and his judgment on the factors he does present will be distorted or uncertain because of the faulty framework against which he sees them.

The task of the grader in evaluating a series of such papers is obviously very difficult. He must visualize the entire framework of analysis and judge all the factors correctly himself and then evaluate the significance of the examinee's approximation to this framework, and his judgments of the separate factors in the light of his imperfect analysis. If, to inexhaustible energy and patience we can add near

omniscience to our grader's qualities, and give him plenty of time, we may get a set of papers correctly evaluated under such conditions. With any lack of these conditions we can expect only the roughest sort of grading.

If the purpose of the examination were to set up a background for further lengthy discussion between the teacher-grader and the pupil such an examination might be justified. But that puts us back in the Hopkins-log-student tradition again, and far from conditions of modern reality. In so far as the purpose of the examination is to secure an appraisal of the examinee's learning achievement, such an examination is very fallible.

The foregoing analysis of the shortcomings of the old type essay-problem examination points the way toward improvement. Since the principal difficulty lies in the fact that the examinee doesn't even come to a judgment of many of the important factors in a problem because of his inability to break down the composite mass, the solution seems to be to take off of his hands a part of the analytical task and to examine him separately upon the several more restricted sections of the original large problem.

Notice, please, that this recommendation is not to relieve the student of all responsibility for analysis, but simply to break down the original composite problem to such a point that its analytical aspects do fall within the examinee's capacity. Only by so doing can we be quite sure that he will not miss *both* the analysis and the evaluation of the factors. And only by so doing can we make it possible for the less than omniscient grader to pass proper judgment on the examinee's performance.

So far the plea of this paper has been for greater objectivity within the general realm of the older type problem-essay examination. It should be obvious that following this trend to its logical conclusion gives us the so-called fully objective type. How, then, does this type of question fit into the scheme of learning-teaching-examining as here outlined?

In the fully objective type the *formally* analytical part of the work has been reduced to a minimum in order to limit the examinee to a single judgment at a time. The framework behind the question is so laid out that there can be no doubt as to what the point of the question is, nor can there be any doubt as to whether the answer given is right or wrong. The question should leave no room for assumptions that might vary the answer; there should be no ambiguity of terms.

All of the limitations in the preceding paragraph do not mean that the question is easy. It may be quite easy in the sense that it requires a judgment as to the effect of one factor only, or it may be very hard in the sense that it requires a judgment as to the joint effect of several or many factors. In every case, however, the given information must be complete as to what those factors are. The factors must be stated with such a degree of clarity that experts in the subject could not possibly disagree as to the answer.

The point which greatly worries a lot of examiners is the fear that this ultimate step in the objectivization of the examining technique finally loses something essential, and this something essential is what is generally in mind when we use the expression "analytical ability".

In the face of a concept of this degree of semantic complexity it may be that only a long series of actual examples can possibly tie down the real meaning of our worry and give us any chance for a meeting of minds. Such a series of suggested examination problems in the field of accounting forms the final section of this brochure. Perhaps, however, it will do no harm to attempt first a brief direct attack on the problem in the form of exposition, which, if it does not prove the case in itself, may at least act as a guide to the later conducted tour through actual examples.

The analysis of a complex problem in accounting is fundamentally parallel to the analysis of a complex chemical compound. While certain routines or rules of procedure might seem to aid in the formal analysis, real success must lie in a clear understanding of what the component elements of the compound or problem might be. Basic to such understanding is not only the nomenclature and description of the elements but an awareness of their relationships and interactions under an immense variety of circumstances.

In chemistry the relations and interactions must be described in terms of color, form, weight, expansion, contraction, stability, hardness, odor and other such physical terms. In accounting the relationships and interactions must be described in terms of significance toward or directional effect upon the basic concepts of accounting, asset, liability, capital, income, expense, accumulation, distribution, proration, measurement, greater than, equal to, less than, favorable, conservative, in agreement with, allowed by, and so forth. In addition to these directional criteria, a very great number of fundamental accounting facts and forces can be measured specifically, under properly described circumstances, in terms of dollars and cents.

Against this background the case for the objective examination in the field of accounting can be stated thus:

> Virtually every problem in accounting can be broken down for testing purposes into its essential elements and relationships, and the understanding of these essential elements and relationships can be tested with such certainty that the examinee's ability to handle them singly or in any combination can be determined with a high degree of reliability.

Perhaps at this point it may be desirable to state again that there is no intention here to argue that objective examinations should replace the older essay-problem types entirely. There will probably always be circumstances in which these older types may be used quite effectively. It should be clear, however, that the circumstances of the present Committee on Selection of Personnel examining project are preeminently those in which the older examination technique would be most seriously handicapped, and which therefore require adherence to the purely objective type. The imperative need for objective grading on a nation wide uniform basis would be determining even if the supposed disadvantages of the objective type for important aspects of accounting achievement could not be so largely dissolved by close study.

Analysis of Subject Matter and the Attainment of Objectivity

The following sections of this brochure have a two-fold purpose: first to illustrate and support some of the contentions of the earlier section in favor of greater objectivity in examinations in general, and second to demonstrate the technique of breaking down some accounting problems into fully objective forms.

Let us assume that we are testing a fairly advanced group and that the area to be covered is inventories. Mark Hopkins might have handled this situation by settling comfortably on his end of the log and saying, "Well, son, tell me what you know about inventories this morning." Many an examination has been given in about this form but under less propitious circumstances. It is literally true that instructors have given examinations by writing on the blackboard, "Discuss the problem of inventories - time 60 minutes."

If the examinees are a group of students who have just finished a couple of days on a certain chapter in the same specific text book, the question isn't quite as vague as it seems. What is wanted is an ample rehash of that chapter with especial attention to the particular instructor's pet additions to or criticisms thereof. If it is known that the instructor will do his own grading, a bright boy should be able to earn himself a fair grade. Here we would still be close to certain of the less attrative aspects of the mossy log school of learning.

If on the other hand the same question were put to a group of examinees who had been trained by different schools, instructors and text books, the papers produced would vary greatly in area covered. The subject is just too big for a single hour's test. The task even of outlining such a subject adequately would be worthy of a text book writer's prayerful labor of many hours or days.

Probably no intelligent test maker would use this question for the conditions stated. By its very exaggeration it makes its point. However, it is the writer's contention that many an examiner is guilty of exactly the same crime, the only difference being one of degree.

Suppose we narrow the question down a degree or two and ask our examinees to discuss perpetual inventories, or physical inventories, or manufacturing inventories. Is the question fundamentally fairer? We are moving in the right direction, but we still have a long way to go.

Suppose we break down the still vague concept 'physical inventories'. We might ask for a discussion on: the routines and precautions to be observed in the actual taking of the physical inventories; the problem of inclusions and exclusions; the problem of the relation of physical to book inventory; the problem of inventorying goods in process; the problem of pricing; and so forth. On a straight "discuss" basis any one of these topics is still too vague for a widely used examination.

Let us break down the general area of the pricing of "physical" inventories. We might ask for discussion, definition or description of the 'methods of pricing inventories'; the effect of higher or lower pricing of

beginning and ending inventories upon the showing of profits; or the peculiar difficulties of pricing inventories of goods in process or finished goods in the manufacturing concern.

Each of the different 'methods' of pricing is itself good for discussion, each could be developed by requiring discussion of its application under varying circumstances which are known to raise questions of interpretation. A very fruitful type of question, whether used in essay form or cast into practical dollar and cents form, is that which calls for comparisons of the results of the different methods under the varying circumstances of price movement, volume change, in-and-out quantity relations, and so forth.

The successive breakdowns suggested in the last few paragraphs are essentially the *topic* breakdowns of an outline on inventories. At any stage of these topics and sub-topics there are usually *aspect* variations which lend themselves to examining. For instance our mention of pricing methods above concerned itself with description and effect of the several methods. If we were to ask for the theoretic justification of the several methods, their acceptability for income tax purposes, their probable managerial utility, the difficulty or cost of their application under various circumstances, their applicability to perpetual inventory procedure - these the writer considers *aspect* variations as distinguished from *topic* variations.

Whatever the general area to be examined may be, the construction of an exhaustive topic and aspect outline will always be found a useful preliminary to test construction. Not only will it uncover questions which might otherwise be neglected, but it will materially aid in judging their real complexities. Some topics will break down to their useful elements easily, as one can shatter a lump of channel coal with a sledge, for others the scalpel of imagination must be carefully wielded to get at the real elements.

Even if we were to confine ourselves to the rudimentary essay type of question calling upon the examinee simply to 'discuss' the series of topics as outlined, the successive narrowing of the subject would tend toward greater objectivity.

The foregoing successive break-downs of the general subject, inventories, do not pretend to be complete. The intent here was merely to illustrate the idea of gradually narrowing a subject down to manageable proportions, i.e., manageable by the examiner as well as by the examinee. The initial stages in the construction of such an exhaustive inventory of any area are not unlike the construction of a text-book table of contents. However, to be really effective as an aid in test construction such an outline should go much further than any normal table of contents. It should become in effect an inventory of all useful knowledge at each point of the normal topic breakdown.

Special attention should be called to the final detailed items which the writer has called *aspect variations*. Skill in visualizing these elaborations of the usual topic outline is of utmost importance in framing good test questions, because it is by the use of these variations, most of which go beyond the straight declaratory statements of the text book or lecture,

that we get into the best testing technique. It is precisely at this point that we get beyond asking for the "parroting back" of memorized material. Here we can require the student to demonstrate his knowledge by using it under conditions which are new but clearly delimited and unambiguously stated.

This type of question has the great advantage of closely paralleling conditions of actual accounting work. Circumstances "on the job" always seem to be just a little different from those provided by the good smooth text book; one is always getting into things from a different angle, always having to take some unusual condition into account, always worried about some sticky practical effect.

Such aspect variations are so divergent, so pleochromatic, as to defy systematic cataloging. However, a check list, though obviously random and incomplete, is here given to suggest the technique of discovering question possibilities. These aspect items should be thought of as applied one by one to the detailed normal outline topic which is the subject of testing. Each check item in effect poses a question: Is this concept of any use in connection with this topic? Many check items will obviously not apply at all to many topics, but the sum total of new ideas that can be hatched out of this nest of eggs is very considerable if the test-maker has a fair imagination and lets it brood for a while.

CHECK LIST OF "ASPECT VARIATIONS"

Applicability
Alternate methods
Comparison
Interpretations
Extension
Recognition
Control
Significance
Chronology
Frequency, relative, absolute
Relations to other topics
Accounting cycle
Business cycle
Effect of proper use
Effect of improper use or error
Accrual basis
Varying schools of thought
Budget significance
Theoretic and (or, versus) practical aspects
Justification for
Contradictions, apparent or real
Concerns what parties, how, (owners, creditors, government, tax authorities)
Size of business
Types of business
Violations of principle, consequences
Reverse conditions, timing
Reverse approach (information-wise)
Authorities say

Common misinterpretations
Vouchering of
Auditing of
Balance sheet significance
Profit and loss statement significance
Organization chart
Chart of accounts
Ledger showing
Books of original entry
Legality
Economics of
Social consequences
Relation to basic definitions

Accounting Tests and Objectivity

Up to this point our discussion of the importance of objectivity and of the techniques of obtaining it have been posed against a background of the essay question. This was done with the intent to bring out more clearly the relationships of the examining problem and the learning process itself. All that has been said is applicable to the use of the essay question in accounting as well as in other fields.

In the field of accounting, however, we are fortunate to have a tremendous variety of concrete problem situations which can be used for testing as well as learning, which by their very nature get away from many of the difficulties of the old essay type. Once we have learned the fundamental 'language' of accounting we are freed from many of the semantic difficulties which plague the teacher and tester in other fields. The difficulties of terminology, about which accountants are properly worried, can be recognized and easily avoided at most levels of testing. They can in fact be used as the basis of searching questions at the higher levels.

Thus it might appear that the "typical accounting problem", being concrete and avoiding semantic difficulties might be objective enough for reliable examining purposes. To the great good luck of our profession this is true in large measure. The result is that more tests in accounting have, by and large, been better testing devices than the typical tests in many other subjects, even without the conscious pressure for objectivity such as this paper advocates.

As long as the accounting problem is kept simple in the sense that all pertinent factors are clearly expressed and that there is only one right answer, clearly recognizable as such, the problem is objective and can be graded reliably. These conditions are most frequently attained in the more elementary phases of accounting. As the subject matter of accounting becomes more difficult, testing tends to lose its real objectivity by the unwise and unnecessary complication of the test problems, until in many instances they represent as bad a burden of analysis as the essay type at its worst. There are, of course, circumstances in which it is important to test for this very analytical ability, and these cases must be taken care of, but the fact remains that up and down the whole range of examining we have too often let the complication of factors pose analysis difficulties so formidable that our appraisal of the examinee's judgments on fundamental relationships has

been badly obscured.

Again it becomes necessary to consider the tactical relationship of the examiner and the examinee. If we are talking of the examinations which Professor Hopkins, himself, makes up, administers, and grades, for 25 or 30 students whom he has had in class for several months, and if Professor Hopkins is quite wise and very conscientious in his grading, his resulting appraisal of the students will be quite acceptable.

As conditions move away from this ideal situation, however, even the relatively concrete accounting problem will not prove a reliable testing device. Examinees with different backgrounds will put varying interpretations on the given "facts" and figures and varying emphasis on apparent requirements. Even greater variation results from the grader's difficulty in evaluating the significance of various deviations from the standard solution. Thus the fact that significant areas of accounting lie in the realm of figures, forms, and standardized terminology, does not of itself guarantee that tests in this field will be objective. We need not search far to find concrete figure problems on accounting examinations which are just as bad fundamentally as anything which the sciences or humanities have produced. The nature of the subject matter of accounting does, however, lend itself beautifully to full objective treatment. This fact has been brought out convincingly in the experimental examinations which the Committee on Selection of Personnel has used to date on several thousands of students and professional people in accounting work. It now seems probable that with the accumulation of experience we will be able to set up fully objective examinations which, for significant areas of accounting achievement, will be at least as valid and reliable as those in the sciences, law, medicine, engineering and other fields in which they are a proven success.

The writer is well aware of the suspicion among many teachers and accountants that the fully objective forms of tests can at best test only "factual" or "memory" items. Such items are the obvious ones which always suggest themselves to beginners in test construction. Probably there is only one good way to overcome this natural prejudice, and that is by setting up a few good samples and letting the reader judge for himself whether the tests can be really objective and still go into worth while levels of accounting achievement.

The Construction of Objective Tests in Accounting

This section has several related purposes all of which we will attempt to serve by discussing several objective accounting problems at different levels of difficulty. It is hoped that this will make a good beginning in convincing the doubtful of the usefulness of this type of test. A final purpose is to show enough of the technique of objective test construction to encourage more accounting teachers to try the technique themselves, in order that with a much wider experience base the examinations of the Committee may be improved in time to come.

To many people the term "Objective Type exam" means just one thing -- "True-False". Others are also familiar with the "completion" (or simple recall) and the multiple choice forms. These three are the backbone of

objective testing. Other types are really variants of these three. A brief look at the advantages and disadvantages of these three forms will be a good background for judgment of any objective form.

COMPLETION OR SIMPLE RECALL

This is the type in which a sentence is set up with a blank space indicating the omission of a word or simple phrase. The examinee is required to fill in the missing word. This technique can be used for straight factual matter: historical dates, places, persons, language forms, mathematical data, symbols, and so forth. Wherever accurate definitions are of interest the form may be used by stating the definition with the thing defined left blank. This type does not appear particularly useful in accounting, and will not be used on the C.S.P. examinations for the very good reason that it does not lend itself to machine grading. It is becoming passe' in many other fields for the principal reason that other forms accomplish all that this form accomplishes and are not subject to some of its tricky disadvantages. Experimentation with the completion type usually develops one discovery -- the perfectly fiendish ability of examinees to discover answer words which are correct in some stretched sense which never could occur to the test maker.

TRUE - FALSE

This form has been widely used and roundly cursed. When carefully used under the proper circumstances it is an effective testing device. It's apparent simplicity and ease of construction have, however, caused it to be used in many instances where it did not fit properly, and this has brought the form into ill repute in many peoples' minds.

The rules for the proper use of the True - False form are simple enough. This form may be used wherever, in the material to be tested, there exist a lot of specific questions which are definitely worth testing and which by their nature are clearly "open and shut" propositions. The difficulty is that in many fields of examining these conditions simply are not met. On one hand the test maker, trying hard to get fifty or one hundred items, takes in a lot of items that are really true or false but too trivial to be worth testing.

On the other hand the test maker, striving for worth whileness, gets into items which are not absolutely true or absolutely false. Within the context of the examiner's immediate thinking a proposition may be true, but the examinee with perfect propriety sees other factors behind the statement which may make it false. At times the very best student will fare worse on certain T - F items than his less able competitor, because his very brightness causes him to go beneath the surface intent of the question to dig up hidden meanings resulting in the answer 'False', when the examiner intended the answer 'True'. Such a set-up is obviously vicious.

Recently a friend of the writer's used the following T - F item on a biology test. "All spadefoot toads of Oklahoma lay large-yolked eggs." It was intended to be answered "true". One student answered "false" and supported his point with the untouchable argument that the male spadefoot didn't lay any eggs at all. The item was later changed to read, "The spadefoot variety of toads lays large-yolked eggs." Another bright boy marked this

false on the ground that it isn't the *variety* but the individual female that lays the eggs. It wouldn't be safe to word the item, "All female spadefoots lay large-yolked eggs", because many females are too young to lay eggs. Possibly it would be safe to say, "The eggs of the spadefoot toad are characteristically large-yolked."

One simple, but very important, caution on construction of T - F items is the observation that such absolute words as all, always, never, and so forth are dangerous. In most areas these absolute expressions make a statement false. Thus either they are dead give-aways if the item is intended as false, or else they give the examinee grounds for argument if his answer of "false" is marked incorrect. The use of the word "characteristically" in the final item in the preceding paragraph illustrates the avoidance of this danger. If the statement had claimed that the eggs were *always* large-yolked, some bright boy would be sure to have found a few spadefoot eggs that for some unaccountable reason had tiny yolks, and it would be hard to prove to him that he was wrong if he marked the item false.

Usually a really critical review of a suggested list of T - F items by one's colleagues should catch the worst "boners", if said colleagues have been properly indoctrinated in the dangers of the game.

Just as professional testers are tending to get away from the use of the simple recall type, so too, many are getting away from the T - F because of the difficulty of its use and because other types accomplish the same purposes with greater safety. The fundamental difficulty of the T - F form in most fields is the rarity of the really open-shut, right-wrong, true-false situations. Technically such a situation is known as a dichotomy (adjective dichotomous, meaning cut in two).

It so happens that while true dichotomies may be rare in other fields they are extremely common in the field of accounting. The discovery of this fact has been one of the most useful developments in the C. S. P. test construction program. At any level of accounting where we are concerned with debit-credit analysis we have true dichotomies. Whether we are asking a beginner how a sale return affects the customer's account or quizzing the C.P.A. candidate as to the effect on capital surplus of an involved reorganization, the answer is specifically debit or credit and it cannot be something in between, nor, *if the conditions are properly stated*, can there be any argument as to which answer is correct. In the area of accounting systems, whether we are concerned with a simple bookkeeping set or with the whole accounting system of the United States Navy a vast number of significant questions can be visualized as simple dichotomies. For example: is the total of the general column in the cash receipts book posted to the general ledger or is it not? Does the total of the Supplies and Accounts Form #222 of the U. S. Navy, as sent in monthly by the Fiscal Officer of a Naval Shipyard, affect the balance of the Naval Stock Fund account in the appropriation ledger at the Bureau of Supplies and Accounts in Washington, or does it not? These are true dichotomies and in both cases the information called for is not a mere memory fact, but a reflection of fundamental understanding of the relationships involved.

It should be noted that these items could be set up as declarative statements and true-false judgments called for, or the same effect could be obtained by setting up series of such questions (within each system) in

tabular form with two columns, "posted, not posted" or "affects, does not affect" to be checked. Since there is a 50-50 chance of guessing the dichotomy right, it is always desirable to set up groups of related questions for the double purpose of getting real coverage and to minimize the effect of guessing.

MULTIPLE CHOICE

Of the three basic types of objective problems the multiple choice is easily the present day favorite in most testing circles. It is widely adaptable as to subject matter. Though it may be used with any number of choices, the five choice form is the favorite. The fact that this form reduces the chance of guessing to one in five is only one of its concrete advantages. Its basic form is a statement with five conclusions, one of which is to be chosen. The choices may be figure amounts, single words, extended phrases, or whole sentences. The general directive for the multiple choice section of an examination may use phraseology which requires the choices to be exactly correct, or it may call for the "best" choice without the handicapping restriction that it must be absolutely and completely *right*. This freedom of the "best answer" form is a great boon in vast areas like the social sciences where the careful scientist hesitates to claim exact knowledge or to sponsor absolute judgments.

When the answer to a multiple choice problem is a dollar amount or other figure the five choices may be expressed as specific figures or as ranges within one of which the correct answer will lie. This adaptation is often necessary to prevent solution by checking backward from the answer which would often be easier than to work out the problem as intended. Such ranges should usually be continuous and equal, with their limits on even amounts to permit ready inspection. They may be open ended. Sometimes the fifth choice may be "some other amount" or even "amount cannot be calculated from the data given." These options may sound mean but in the proper places they are effective.

In any multiple choice problem whether the answer is to be a word, phrase or amount, care should be exercised to see that the incorrect choices, or decoys, are as plausible as possible. If one or more of the decoys are patently wrong it simply means that our supposed five-part problem has degenerated into a four-, or three-, or two-part problem with a resulting increase in the guessing chance. The ideal procedure for determining effective decoys is to give the problems in free-answer form to a representative group of examinees prior to setting up the multiple choice examination. If the free answers show any concentration of errors such errors will presumably make the best decoys. If this procedure is not available, the experienced teacher will do well to rack his brain carefully for potential errors, for the validity of this type of problem increases greatly if the decoys are well chosen.

We have said that five seems to be the optimum number of choices, but this should not be taken to call for the forcing of five choices into a situation which naturally calls for less, and it so happens that accounting is full of situations that make natural three-choice questions. A great many real accounting problems concern the effect of some process upon resultant figures of gross profit, net profit, surplus, current assets, total assets, current ratios, turnover, and so forth. Whether the process in

question would increase, decrease, or not affect each of these resultant figures is often the essence of the debate. Such situations are naturals for three-part multiple choice problems. The particular form in which to throw the problem depends upon the exact point it is wished to emphasize. Two processes can be described and the question put in the form: Will the effect of process A be to show a greater, lesser, or equal net profit to that shown by process B? A series of suggested changes can be described and the querry be whether each would produce a net profit greater than, less than, or equal to the net profit of the present process. The question can be put in terms of the neglect or omission of certain assumed proper processes and the effect thereof upon net profit as being too great, too small, or unaffected. Anywhere in the field of accounting where methods are to be compared, where changes are contemplated, or where the significance of errors is to be judged, the three-part multiple choice problem may be the most effective testing device.

MATCHING, CLASSIFICATION, TRANSACTION ANALYSIS

In objective tests in elementary and high school subjects one frequently comes across what are known as *matching* tests. In a history test for instance one might find the matching test used in this way. Two parallel lists contain the names of famous generals and famous battles. The items are numbered or lettered, and the task of the examinee is to indicate which generals fought which battles. This can be done by putting the number of the battle in the parenthesis after the general's name or by some other simple indicative system. If the battles and generals just come out even, and if the directive says that there is a one-for-one relationship we have a simple task. If there are more battles than generals, or if several generals cooperated in some of the battles, it is easy to see that the task can be made more difficult. If we have ten generals and fifteen battles we have in effect ten fifteen-part multiple choice problems, or by reversing the instructions fifteen ten-part problems. Is this type of thing adaptable to accounting?

A direct use of this matching technique in accounting would be to set up two lists, one comprising 25 accounts, and the other 10 classifications such as: current asset, fixed asset, current liability, fixed liability, net worth, operating income, non-operating income, manufacturing expense, selling expense, general administration expense. The directive would require the accounts to be identified as to classification. In this form the problem would not be very effective since *good* account titles would make the problem too easy, and we are hardly justified in using *poor* titles just to make the problem hard.

A usable adaptation of the above matching problem would be as follows: Use the five balance sheet classifications only. Instead of 25 account titles to be classified, write descriptions of 25 "objects, facts or circumstances" to be considered in setting up a balance sheet and require the examinee to identify each with the section of the balance sheet in which it would appear. In this form the problem can be made as easy or difficult as desired; it can get away from standardized terms; and it can test the understanding of really fundamental concepts. Having five classifications this form really becomes a five-part multiple choice with the incidental practical advantage of fitting nicely into a machine grading scheme.

An appreciation of the relations, variations and peculiar capabilities of some of these more or less standardized objective test forms is very useful in attacking the problem of reducing any new test area to manageable form. A good illustration is our old favorite examination topic of bank reconsiliations.

After some experimentation the writer discovered that a bank reconciliation makes a natural four part multiple choice problem. There are two classes of items to be considered in a bank reconciliation: those which the depositor has recorded but which the bank has not recorded; and those which the bank has recorded but which the depositor has not recorded. Each of these breaks into two divisions, being either an increase or a decrease in the balance. This gives us *four* technical groups of items. If these groups are carefully defined, against the background of a reconciliation form, we have the framework for our four-part multiple choice problem. We can then proceed to describe as many items as we wish, making them as easy or as complicated as the situation requires.

Another apparently useful and very flexible adaptation stems from our old fashioned elementary problem: "Journalize the following transactions." Such a free answer form would not satisfy objective standards because there would be so many not-quite-right and special-case-right solutions caused by the entirely free choice of account titles. A first step in tying the journalizing problem down to objective form is to set up a chart of accounts and to require the journal entries to be made up from these accounts only. This will eliminate the worst vagaries in the solution. The actual form of expression can be simplified by requiring the student to express his answers by giving the *numbers* of the accounts debited and credited instead of by writing out their titles. If there were 50 accounts in the chart, each debit and each credit required would be a 50-part multiple choice problem. So many choices give us virtually a free answer form of test.

To bring this test further under control, i.e., to make it convenient for machine grading, it can be put into 5-part multiple choice. This is done by selecting five account numbers from the chart for each debit and each credit and requiring the examinee to make his choice from these five numbers. The assumption here is that the test-maker could pick the four most likely decoys to constitute, with the correct answer, the five choices offered.

To try out the test-maker's ability under this assumption one of the C.S.P. experimental tests was run in two forms: the first as a free-answer form (limited only by the chart of accounts), and the second a five choice form made up as described above. Several hundreds of students in each of two presumably equivalent groups took these tests and a detailed comparison was made of the free wrong answers with the set-up decoys. This experiment taught the test-maker (who happened to be the writer) some interesting things about the way the student mind works. Roughly the set-up decoys were the ones the error-makers wanted, but the differences were enough to keep the test-maker humble.

One very common error not clearly foreseen was the examinee's passion for picking the right accounts and then reversing the debit and credit. Another common error can be described as literal-mindedness. If the de-

scription of the transaction contained the words "note payable" a certain number of examinees seem to insist on debiting or crediting the note payable account, even though the transaction had to do only with the adjustment for interest accrued on a note payable. These two discoveries would indicate that in future on such problems the student should be given every opportunity to make these errors. This may sound like scientific entrapment, or a plain dirty trick, but it is also effective testing.

Objective Testing at Higher Levels

The study of older testing programs in other fields which have accumulated a great deal of experience teaches us that a really good test is effective over a considerable range of ability and training. If we want to test the real ability and achievement of a man whose training has gone through the successive stages A, B, C, D, and E we should not confine our test to the subject matter of stage E, but will get more reliable results if we use the best possible testing materials from all five stages.

If stages A, B, C, D, and E were college courses, the instructor of course E would naturally test his students primarily on the subject matter of course E. The problem confronting the Committee on Selection of Personnel, however, is to test accounting ability and achievement, at whatever presumed level, by the most effective tests available. Implementation of our testing principle to serve the purposes of the Committee will therefore mean that there will be no radical change of form or aspect of its examinations from one level to the next. The same principle assures us that there will be no need, at even the higher levels, to drag in specialized subject matter from specific businesses or unusual circumstances. In other words, there is enough good testing material right on the main line of accounting progress to find out who is who in accounting at any level.

The fact that we can make this statement should be reassuring alike to the prospective examinees and to those responsible for their training. At the college levels (I and II) the content of courses typically taken by the accounting student will cover the subject matter quite adequately. At level III the experience of virtually any junior or semi-senior in typical public accounting practice plus such study and reading as he would normally do will cover the additional subject matter required for the C.S.P. test.

With this general statement as a background it would be desirable to attempt to give some indication of what can be done within the framework of objective test requirements to advance the difficulty by properly controlled steps.

This task requires a short back-flash to what was said earlier in this paper in the discussion of test materials. Certainly we do not now want to throw back at the student the hazy and hazardous task of analysis which was so vehemently disclaimed before. The examinee must be under no doubt as to what the question is which he is required to answer.

If our tests are to remain objective then every problem must be such that its answer can be expressed by number, check mark, or the selection of one of the given choices. Every problem must furthermore be objective

as to the pattern of thought which leads up to the certain answer. The objective pattern of thought must mean that the factors bearing on the decision must be unmistakably presented so that there could be no difference of opinion among experts as to the answer.

At the most elementary level the factors are few and simply expressed. As the problems grow more difficult, more factors are thrown into the picture. *This is the essence of growing difficulty -- more factors.* These factors have to be considered, held in mind, weighed, related to one another, and finally combined into a net resultant. This is the fundamental process of thinking in the field of accounting, and it is the ability to perform this process for which we are testing.

Closely related to the problem of weighing and combining numerous factors is the ability to *recognize factors* as relevant. We have a very close parallel to this situation in the elementary algebra courses which we all remember more or less hazily. For each general area in the course the algebra text customarily presented a lot of formal equations of the particular order to be studied, and our task was to "solve each one for x." Then a little later we were required to work the "written problems" which were ostensible descriptions of real life situations. Here our task was to ferret out or recognize the factors and to set up the equation, which could then be solved by our previously developed methods. In our accounting tests there will not be quite the sudden break from "equations all set up" to "written problems", but the harder tests will contain more and more of the element of real life situations from which the examinee must ferret out the factors which he must use in coming to his conclusion.

Lest even this sound like going back to the old hazy analysis problem of the type "Discuss goodwill", it should be pointed out that there is to be no doubt in even our most difficult tests as to what the problem is. There may be a dozen factors hidden in a complicated description of a situation, but the question to be answered will be entirely specific, for instance: At what cost shall the new machine be booked? What is the net effect on surplus? Does the auditor discover the fraud? Should the company make or buy?

Panorama and Detail Technique

Just how, as a matter of mechanics, can the higher level test be set up to satisfy the conditions of difficulty and objectivity as above outlined and still stay within a decent length time-wise and wordage-wise? Remember that we would like to have 100 or 150 independent answers in our test. It would take a lot of wordage to set up 100 five-part multiple choice problems each containing enough factors well disguised for realism and still have each problem really independent of the other 99.

The device which seems to be the answer is the panorama and detail problem technique, which works like this:

A general background is described or otherwise presented in considerable detail. This background may be a whole business situation; it may be the whole accounting system of a business; it may be a part of that accounting system such as its accounts receivable procedure, or its payroll

routine; it may be some general management problem; it may be a lengthy authoritative quotation; or it may be any other general situation which a whole series of specific problems can be posed. Each of the specific problems may present additional specific information which is to be considered against the general background in the solution of some particular point. This technique provides amply for a testing of the examinee's ability to sort out the significant from the inconsequential and to combine all pertinent factors into a sensible answer. And still the final answers may be such that a dozen experts could not possibly disagree on the answer.

The criticism has sometimes been made to objective examinations that they are not realistic. This criticism does not seem valid against the panorama technique. This technique seems to parallel conditions of actual accounting work quite closely. Against the background of the business for which you are working, or the business under audit, you are required to make a dozen or a hundred specific decisions, and every decision must respect the background factors as well as the data immediately in hand.

Validating the Objective Examination

One of the very great practical advantages of the objective type examination is its susceptibility to the process known as validating, or item analysis. Consistently applied, this process is an effective device for constantly improving any system of such examinations.

The ideal situation is that in which the objective test can be tried out on a sizeable and representative group of examinees for the sole purpose of testing the test before it is put into serious use in testing the examinees. Fortunately the Committee on Selection of Personnel has been in a position to do just that, through the kind cooperation of a number of colleges and several large firms of public accountants.

Experience proves over and over again that even the most carefully constructed test will contain items which vary widely in difficulty and in their ability to discriminate between the sheep and the goats. This discriminating ability of a test item or problem is known technically as "validity". Occasionally items which looked perfectly good to the test-maker turn out to have very low, and sometimes even negative, validity. The discovery and correction or dropping of such items strengthens the entire test very materially. The validation procedure works about as follows:

Let us say we have tried out an experimental form of an examination on 350 students. The graded papers are sorted in order of the grades; the middle fifty percent are laid aside, the high and low quarters being the useful ones for validating purposes.[1] Now a study is made of each item in the test showing just how many students in each of the two groups succeeded in answering it correctly. If an item has good discriminatory value it should be answered correctly by many more of the high group than of the low group. The accompanying form is a useful one for reporting

[1] Theoretically the high and low 27 percent give the best discrimination, and these percentages are usually chosen, although actually the validities will not vary appreciably when high and low quarters are used instead.

such an analysis - in this case it is set up for a five-part multiple choice examination.

ITEM ANALYSIS

TEST: (ADAPTED)

	HIGH GROUP								N = 87	LOW GROUP								N = 87	
No.	1	2	3	4	5	0	D_H	CD_H		1	2	3	4	5	0	D_L	CD_L	D	V
1	0	1	(83)	0	3	0	83	95		3	7	(55)	0	22	0	55	63	79	49
2	2	(65)	10	8	2	0	65	75		12	(37)	14	21	3	0	37	43	59	34
3	3	0	(66)	1	17	0	66	76		25	10	(37)	2	12	1	37	42	59	36
4	(53)	17	13	2	2	0	53	61		(50)	1	2	28	6	0	50	57	59	04
5	7	9	(53)	3	15	0	53	61		10	11	(37)	21	8	0	37	42	51	19
6	0	0	0	0	(87)	0	87	100		0	0	1	11	(75)	0	75	86	93	46
7	2	0	1	(84)	0	0	84	96		5	3	0	(73)	6	0	73	84	90	30
8	10	26	27	(23)	1	0	23	26		24	6	2	(32)	22	1	32	37	32	-10
9	(58)	0	2	8	19	0	58	67		(44)	0	18	23	2	0	44	51	59	17
10	0	(83)	1	1	2	0	83	95		1	(43)	3	18	22	0	43	50	73	58
11	30	15	4	16	(22)	0	22	25		36	29	4	13	(5)	0	5	06	15	34

The first column at the left gives the number of the item. The numbers in the columns headed 1, 2, 3, 4, 5, and O indicate the frequencies with which the examinees picked the choices as numbered in the column heads; numbers in the "O" column mean that the examinee omitted the item entirely. The circles indicate the correct choices. Thus in item 1 we see that 83 out of the high group of 87 picked choice #3 which was the correct answer. In the low group 55 out of the 87 picked choice #3. The rest of the numbers give a clear picture of the relative attractiveness of the decoys. It would appear that choice #4 was an ineffective decoy since not one of these 174 examinees selected it.

The D_H and D_L columns record the number right in the high and low groups respectively. The CD_H and CD_L columns convert the number of rights into the percentage of rights. Thus the 83 rights in the DH column becomes 95 per cent right in the CD_H column.

Item #1 was right in 95 per cent of the high group and 63 per cent in the low group which gives an average for the two groups of 79 per cent shown in the D column. This D column gives a good picture of the relative difficulties of the items. In the illustration they ran from 15 per cent to 93 per cent right for the two groups together.

The final column headed "V" is the index of validity. The calculation of V is pretty technical mathematics in the realm of bi-serial correlation. The general pattern however can be clearly seen by inspection. Fundamentally the validity is a measure of the *spread* between the *percentages* right in

the high and low groups. It does, however, involve another factor as will be seen by comparing items 1 and 2. Each of these had a spread of 32 points between its high and low group right percentages. (95-63 = 32; 75-43 = 32) but the validities are quite different, item 1 being V = 49 and item 2 being V = 34. This gives effect to the principle that a spread of 32 points on a very easy or very hard item is obviously more selective than the same spread on a question of medium difficulty. Comparison of items 2 and 11 brings out the same principle in a different way. Though item 11 had much smaller spread than item 2 (i.e. 19 as against 32) its validity was just as high (34). In the same manner a relatively small spread on a very easy question would give a high validity. In practice the validities are read from a prepared table (Flanagan's) in which the percentages of successes of the high and low groups are arranged along the horizontal and vertical margins of a page and the validities are read at the intersections of lines projected from the horizontal and vertical readings.

In passing final judgment on test items which have been analyzed in this manner the usual rule is that the difficulties should range between 20 and 80, though it is usually desirable to retain a few easier items near the beginning of the test, for their psychological value in getting nervous individuals off to a good start and a scattering of harder items to discriminate among the very able individuals. Validities of 40 or above are ordinarily regarded as unquestionably useful; those of 30 to 39 as fairly satisfactory; those of 20 to 29 as doubtful; those below 20 as clearly indicative of a need for revision or elimination.

The C.S.P. experimental tests were intentionally longer than the final tests were to be. Actual time consumed in the experimental runs was recorded by sections of the tests to give a basis for timing the final form.

Setting up the final form of each test was thus done with the complete story of timing by sections, and the difficulty and validity of each item available. All items which did not appear entirely desirable were dropped. The result should be a very satisfactory final test for the areas of accounting achievement covered.

Some Sample Validations

After the foregoing generalized discussion on the technique of item analysis, it may be of interest for the reader to see some actual results on tests included in the C.S.P. experimental series. The following data are highly concentrated, omitting the analysis of separate decoys in each item and giving only difficulty and validity indexes for each item.

TEST I-A, LEVEL I, STUDENTS, MEDIAN TIME 27 MINUTES.

ITEMS 1–30 THIRTY ACCOUNTS TO BE CLASSIFIED.
ITEMS 31–40 BOOKKEEPING – ENTERING TRANSACTIONS IN BOOKS.
ITEMS 41–58 POSTING BOOKS TO GENERAL AND SUBSIDIARY LEDGERS.

#	D	V	#	D	V	#	D	V	#	D	V
1	99	11	16	99	17	31	97	16	41	70	72
2	80	35	17	41	44	32	97	16	42	77	72
3	92	50	18	90	19	33	92	37	43	78	70
4	29	29	19	61	28	34	94	43	44	78	70
5	66	18	20	96	19	35	90	53	45	66	73
6	96	37	21	54	51	36	30	30	46	74	73
7	56	37	22	90	53	37	48	14	47	76	72
8	31	40	23	52	44	38	95	28	48	67	74
9	22	21	24	17	10	39	77	49	49	60	72
10	60	34	25	22	03	40	90	43	50	55	61
11	17	37	26	15	24				51	54	78
12	84	39	27	48	41				52	58	70
13	49	28	28	57	54				53	62	68
14	99	17	29	58	58				54	54	67
15	99	17	30	64	36				55	55	80
									56	57	77
									57	65	60
									58	60	72

The first forty items of this test were on the whole too easy, eighteen of them rating 80 or above in the D (Difficulty) column. The validities were also rather poor, only ten of them hitting 40 or above, and twelve being below the 20 point.

Items 41-58 turned out to be far better than the first 30. Every item in this group was satisfactory from both difficulty and validity points of view.

TEST II-G, LEVEL II, PROFESSIONAL GROUP, MEDIAN TIME 35 MINUTES

ITEMS 1–12 COMPARISON OF RESULTS UNDER DIFFERENT METHODS OF INVENTORY PRICING.
ITEMS 13–32 LIFO INVENTORY — RESULTS UNDER VARIOUS CONDITIONS.

#	D	V	#	D	V	#	D	V
1	69	78	13	80	62	25	79	58
2	68	71	14	71	60	26	55	67
3	67	75	15	71	72	27	67	66
4	71	76	16	74	68	28	79	62
5	45	71	17	68	64	29	74	47
6	90	38	18	74	69	30	63	42
7	70	77	19	70	72	31	67	71
8	69	70	20	69	73	32	54	39
9	63	42	21	81	46			
10	69	78	22	73	65			
11	70	69	23	76	67			
12	62	40	24	74	68			

Note that this test was at Level II and was given to a group of practicing public accountants. Only two of the validities show below 40 and these are 38 and 39. The difficulties show the test was too easy for this group. It should have had a sprinkling of items down in the 20s and a fair number of 30s and 40s. This test was originally intended for students, and it may have been right for them, but this particular subject matter seems to have been "duck soup" for most of the practitioners.

TEST II-D, LEVEL II
STUDENTS, MEDIAN TIME 29 MINUTES, 48 SECONDS
PROFESSIONAL GROUP, MEDIAN TIME 30 MINUTES, 30 SECONDS.

FIFTY ITEMS (25 REPORTED HERE), TRACING THE EFFECT OF ERRORS ON NET INCOMES, WORKING CAPITAL, CURRENT RATIO AND MERCHANDISE TURNOVER RATE.

	STUDENTS		PROFESSIONAL	
#	D	V	D	V
A-1	74	73	82	66
2	66	54	73	65
3	70	60	80	68
4	83	39	82	66
5	79	58	79	58
B-1	71	58	67	79
2	80	56	79	69
3	59	53	70	69
4	61	71	68	78
5	52	50	51	30
C-1	70	41	72	75
2	83	38	82	66
3	53	49	69	70
4	52	69	67	72
5	58	41	53	34
Ave.	67	54	72	64

	STUDENTS		PROFESSIONAL	
#	D	V	D	V
I-1	44	19	45	26
2	31	11	38	34
3	21	14	32	42
4	20	(-)03	33	41
5	33	15	44	25
J-1	21	14	32	51
2	34	03	36	44
3	28	02	44	31
4	25	28	39	44
5	74	26	70	48
Ave.	33	13	41	39

Test II-D gives us an interesting chance to compare student results with those of professionals in addition to giving us fundamental criticism of the test items. You will note that in the first fifteen items (sections A, B, and C) the professionals had a slight edge over the students in respect to the per cent right. With few exceptions these fifteen items show higher validity for the professionals than for the students, though the student validities are still very acceptable.

In the last ten items (sections I and J) we see that the difficulties were stepped up considerably for both groups. We note further that the validities of these ten items are still quite good for the professional group, but that they are very poor for the student group. There seems to have been something in these last ten items that was still very useful for testing professionals but which ruined the items as far as the students were concerned. Sidelights like this lift the job of item analysis well out of the realm of dry statistics and give it interest and perhaps real significance in examining technique.

Standardizing of Successive Examinations

In addition to the advantage of improvement through the validating technique, objective tests have another very important advantage. This is the fact that by appropriate experimental procedures, the grades on two or more editions of the tests can be expressed in terms of a scale such that a given numerical score or grade will indicate the same level of ability or achievement on each test even though the tests are given successively, months or years apart.

In circumstances where it is considered desirable to use a substantial portion of essay questions, the objective part of the examination can be used to standardize the entire examination in successive years.

Let us assume, e.g., that a certain Board of Examiners has allotted one-fourth of the total examination time to objective tests whose comparability has been experimentally established and verified. This Board identifies its examinations by the calendar years in which they are administered. In 1937 and 1938 the candidates secured these average scores:

	OBJECTIVE TESTS		ESSAY EXAMINATIONS	
	1937	1938	1937	1938
Average	50	51	57	64
Standard Deviation	20	20	8	9

The 1937 and 1938 candidate groups were about equal in size, came from the same colleges, with the same average grades in the same general and special courses, and were of closely similar age and sex constitution. On the 1937 and 1938 objective tests, which were parallel in content, and equivalent in validity and difficulty as determined by careful experimental procedures followed by the Board, the two groups of candidates secured averages of 50 and 51, and the spread (Standard Deviation) of their scores above and below these averages was the same for both groups (20).

But on the essay part of the examination the average scores were 57 and 64. Under the old rules, with a passing grade of 60, 35 per cent would have "passed" the essay examination in 1937, and 67 per cent in 1938.

Under the new rules of the Board whose recent experience we are here "paraphrasing", however, the reasonable assumption was made that the two groups of candidates were in fact about equal, and that the 1937 essay average of 57 was closely equivalent to the 1938 essay average of 64. Since the public welfare seemed to require about the same number of recruits in 1938 as in 1937, the Board "passed" the same number (within two per cent) in both years.

Whether the technique is used for such a combination type examination or is used for the wholly objective type, the practical advantage of being thus able to standardize can hardly be overestimated. It is exactly the quality which accrediting bodies of all kinds have always sought in the examinations. Properly used, it can give the true comparability and fairness to successive year's examinations which the law-makers and administrative bodies most certainly desire.

The Absolute, the Relative, and Policy Decisions

The public schools I went to in my youth had a standard of 80 as the lowest passing grade. I remember my sense of superiority upon learning that some schools "passed" with grades as low as 75. Many reputable schools and colleges recognize 65 as their lowest passing grade. Certain states have embalmed in their C.P.A. laws the dictum that "70 shall be a passing grade in each one of these subjects."

What is the actual meaning of these expressions? Is it possible in any one of these jurisdictions to define what shall be 100 per cent of the knowledge which is possible of achievement, and then, by an examination, to ascertain of what percentage of that total knowledge an individual is possessed? Surely no examination pretends to examine *all* the knowledge in its field. If it does not comprehend all the knowledge, then surely the examination-maker must presume to great skill in sampling if one per cent or one tenth of one percent of all knowledge which his examination does embrace is truly to represent the entire field! Such an interpretation of the situation gives us a flimsy basis for the important decision of admitting to the company of gentlemen or consigning to the outer darkness!

Fortunately for society and sanity, though examiners have sworn to the absolute, their real judgment has almost always been better in action than their words. Only rarely do we hear of the professor who flunks 60 per cent of his students one year and 10 per cent the next, as would most certainly happen if these absolutes were taken as seriously as they are proclaimed. It is almost always found possible to "adjust" a little here and there until somehow the "final results seem reasonable."

Sometimes the professor-examiner actually feels a certain sense of guilt as he strong-arms another dozen grades up into the passing area, and adjusts the other pockets accordingly. The truth of the matter is, however, that this adjustment of the stated standard is the essence of scientific grading procedure. It is common sense recognition of the impossibility of the supposed absolute numerical standard. The professor thereby recognizes that his concept of the totality of measurable knowledge is hazy, that his sampling was probably faulty, that the pre-determined passing percentage is meaningless, and that the breakfast he had that morning (hamless, eggless and with poor coffee, during the war) probably had pulled all the grades down considerably anyway. He furthermore recognizes that the one reasonably stable factor in the whole picture is that, year after year, the students who pour into his required course are a fairly stable cross section of the population with which he has to deal.

He realizes, also, though perhaps hazily, that in setting the pass-flunk line he is dealing with matters of policy that go far beyond any simple arithmetic. He is aware that schools of the kind he serves get along pretty well if in general about 95 to 98 per cent of the sophomores pass these required courses.

Our college professor's problem in his own class parallels the problem of other examining authorities. Insofar as he and they are wise they will recognize that the problem is two-fold. The first part of the problem is to distribute the individuals in the class or other group along a scale

with regard to ability, aptitude, achievement or other quality at issue. This is the function of the examination, whether it be an old-fashioned essay type or the most streamlined objective type. It is worth repeating that this is *the* function of the examination and it is its *sole* function,- to spread out the examinees along an equal interval scale. One to 10, 0 to 180, or 50 to 93, it doesn't make any difference what the scale is.

Having spread out the examinees so that it becomes clear who is better than whom, and by how much, the second part of the task may be undertaken. The professor's job is to decide where the ax shall fall - where along his impromptu scale he shall draw the line separating the goats from the sheep. The consequence of his decision may be refusal of permission to continue in the course, dismissal from the school, or even ineligibility for football. In the long run his decision involves the maintenance of school standards. It is a difficult, *policy* question, a real burden on the professor's mind and conscience. But it is *not* something that the *examination* can do *for* him.

If we were talking about the problem of a college admissions committee, the important line to be drawn might be near the middle of the group examined or it might embrace only the top 25 per cent. Here again the drawing line is a *policy* matter and there are a large variety of factors to be considered; housing capacity, faculty load, tradition, standards, finances. The examination can rank the applicants, but it cannot answer the sixty-four dollar question.

If we visualize the over-all task of a state board of examiners (for law, medicine, nursing, or any profession) we see the policy part of its problem as an especially delicate matter. The real factors in the decision are, or should be, the need to maintain the highest possible standards of the profession represented by the board, and at the same time to qualify enough new members to serve the public welfare. The limiting factor of a legally fixed *numerical* "passing mark" is a handicap to the board in performing its policy function. We have all heard of examinations which in certain years were unintentionally hard and which, by the rigid pre-determined *numerical* standard, "failed" 60 per cent. The chances are that these examinees were not really a less able group than those who took the examination the previous year when only 30 or 40 per cent "failed." The difference lay in the examination, the grading standard, and the unfortunate insistence on an absolute passing mark.

In cases like this the standard which was intended to be fixed and safe, has actually become unpredictably and uncontrollably erratic. If the board cannot or does not use strong-armed "adjustments" the original purpose of the law is badly distorted. It may mean that in one year the public loses the services of a number of able and well equipped men, and in another year the bar proves so low as to admit persons not qualified in the true sense.

In contrast to the double task which faces the course professor, the committee on admissions, and the state board of examiners, the Committee on Selection of Personnel has only one part of the problem to worry about. Its job is to give a good examination, grade it accurately and rank the examinees. A report will be made to each school on its own students. This report will show the median for the school and the national median. The

report on each student will be in terms of his percentile, that is the proportion of the national group that his own score equals or exceeds. A percentile of 86 would mean that his grade was better than the grades of 85 per cent of the examinees. A percentile of 35 would mean that his grade was better than the grades of the lowest 34 per cent of the examinees, and lower than the grades of the highest 64 per cent.

The problem of doing anything about these grades is clearly not the function of the Committee. The various persons interested may take whatever action seems appropriate. The student who finds himself at the end of his sophomore year in the very low percentiles may feel that he should switch his chosen major from accounting to journalism or geology. There need be no attempt to set a limit even in terms of percentiles to mark off the accounting major prospects from those who should seek fame and fortune elsewhere.

The school which finds the median grade of its students near or above the national median may be satisfied with its job. A great many complicated factors will have to be considered in deciding on any action if its median seems unreasonably low in the national picture, and is again clearly a policy matter for the school administration. Such information - the rank of each particular school in comparison with others - is recognized by those in charge of the project as an extremely delicate matter. It will be held in strictest confidence by the few representatives of the Committee who must know the figures in the process of working them up and reporting them to the schools concerned.

Prospective employers will be authorized by job applicants to obtain their percentile standing from the Committee. It will be entirely up to the employer to decide if he wants to take on the boy with any particular percentile rating. For example, if a firm finds after using the tests for a reasonable time that 80 per cent of employees with percentile scores of 60 or above give satisfactory service, and 90 per cent of those scoring 20 or lower fail to render satisfactory service, it is obvious that, for that firm with its own particular standards and conditions, two important reference points have been established for the guidance of the firm's personnel officer in selecting new employees, and in studying service records for the elimination, retention, or promotion of old employees. It cannot be stressed too emphatically, moreover, that the application of even such critically determined scores, must constantly be checked by the application of other essential factors, not measured by the tests, such as college grades, extra-curricular activities, references and other evidence on character, personality, industry, and dependability.

Note that in the three circumstances described, the examination itself was confined to its proper sphere - the ranking of the individuals examined. The use that the student, school, or employer makes of the grade is clearly a separate function. There is no setting up of any absolute or bogey grade which must be passed. In each case the interested parties use the percentile grade for exactly what it is, a measure of accounting achievement expressed relatively. Whether the particular measure is high enough or not high enough for the purpose at hand is a separate policy judgment.

The earliest stages of technical work on the present project have

involved relatively few persons. This task has comprised mainly the study of the general principles of objective test construction and investigation into the applicability of this technique to the subject matter of accounting. It so happened that the writer had experimented with objective testing in the accounting field for a considerable number of years. The present project, however, gave him the incomparable advantage of the consultation and advice of Dr. Ben D. Wood, director, and Dr. Arthur Traxler, assistant director of the project, who were leading experts in the field of testing, and whose great experience in testing in other technical fields was thus brought to bear on the present task.

The earliest proposed tests were reviewed and criticized by members of the committee, and several other practicing public accountants as well as by a number of teachers of accounting. The resulting improved tests were given a real try-out in the spring of 1946 through the generous co-operation of some thirty colleges and the staffs of several large accounting firms. The need for such testing of the tests has been explained earlier in this paper in the discussion of the validating technique. The resulting improvement in the tests is a measure of the absolutely necessary help which the schools and accountants have given. The writer wishes to express for himself, the project director, and the committee, full appreciation for this invaluable aid.

These earlier phases of the project, the writing, criticizing, and try-out testing must, of course, continue, and it is the sincere hope of the writer that many accountants and teachers of accounting will be moved to try their hands at objective test construction. Such wide-spread interest is necessary in order that the almost infinite possibilities may be more thoroughly explored and that constantly more effective forms of test problems may be made available.

As the project swings hopefully into its final phase it will again lean heavily upon the accounting faculties of the colleges, upon whom will depend the task of administering the tests to their successive groups of students, and of making the best use of the results of the lower level test in the problem of guidance. If this double task is handled in as effective a manner as we have confidence it will be handled, a steady and substantial improvement in accounting training and selection should be assured, which will be worth many times the cost and effort which are being put into it.